

Continuous models for production flows

Dieter Armbruster, Christian Ringhofer,
Department of Mathematics, Arizona State University
Tae-Chang Jo, Mathematics Department
New Mexico Institute of Mining and Technology

Abstract

We model high volume, multi-stage continuous production flow through a re-entrant factory. We approximate production volume through a continuous density variable and production stages through a continuous completion variable. Ideas from traffic modeling and gas dynamics are adapted to generate a hierarchy of PDE models. All of these models are variations of nonlinear nonlocal hyperbolic conservation laws. They allow extremely fast and accurate simulations. Large scale discrete event simulation based on an INTEL factory are presented. The data are analyzed to generate a parameterization for the PDE model of this factory. A diffusion coefficient is determined by looking at the spread of throughput times in a factory run at steady state.

1 Introduction

A fundamental goal of any supply chain simulation is to generate simulation tools that support the exploration of business questions and to pose what if? questions on these simulations. Since most production deals with individual parts and the processes that these parts undergo, the natural method of choice for accurate simulations is discrete event simulators. While they have been very successful on the factory level [11], e.g. to simulate semiconductor production lines, they are relatively slow, and it seems obvious that they are not scalable to a full supply chain. Based on analogies to

traffic flows and using methods from gas dynamics we have therefore recently developed mathematical models of production flows. In a modern semiconductor factory we are interested to model and simulate on the order of 250 production steps executed on about 100 machines, with a re-entrant part of the production line that cycles about 20- 30 times. Hence the design feature of all the new models has been that they are independent of the number of parts and processes in the factory. In fact their performance will be more accurate the larger the number of parts and processes. Another problem associated with discrete event simulators is the fact that they are stochastic simulations. While it may be straightforward, though still time consuming, to determine probability distributions for e.g. throughput of a factory run in steady state by doing many runs, it is much harder to determine meaningful average and statistical results for time dependent system. However, with the life cycle of a typical semiconductor chip of the order of one year and a throughput time between 40 and 60 days it is unlikely that a factory is ever run for any longer amount of time in steady state. We are therefore especially interested in *transient* behavior of such systems and have developed a hierarchy of models that allows the simulation of the time evolution of moments of the probability distributions of the stochastic production process [3], [1],[2]. These simulation models are intended as building blocks for large supply chain models [10] as well as for the prediction stage in Model Predictive Control algorithms [14].

2 Quasi-steady models

In addition to Discrete Event Simulations another common way to model production processes is the use of fluid networks. Fluid models [5], come from traffic theory and were introduced by Newell [13] to approximately solve queuing problems. They consider the length of a queue $q(t)$ as a continuous variable whose rate of change is given by:

$$\frac{dq}{dt} = \begin{cases} \lambda(t) - \mu(t) & \text{for } q(t) \neq 0 \\ 0 & \text{for } q(t) = 0, \end{cases} \quad (1)$$

where $\lambda(t)$ is the arrival rate and $\mu(t)$ the processing rate of the queue. This basic building block for a queue can be connected to a *work-conserving fluid model* by feeding the outflux of each queue into other queues. In terms of aggregate modeling, a fluid network stops half way. While a fluid network models the work in progress with a continuous variable, it models machines as individual discrete queues. With several hundred production steps for a typical chip it is reasonable to approximate the production steps also along a continuum.

We therefore developed a heuristic model for a truly continuous description of the production process (see [3] and [2]): Let x be a continuous variable representing completion of the product, i.e. product at $x = 0$ denotes a raw product and parts at $x = 1$ denote a finished product. There is a unique production process assigned to every x -value. Assuming a high volume, many stage factory we model production flow with a continuum variable on a continuum domain. We write $u(x, t)$ for the (WIP) density of product at stage x and at time t . Assuming a unique entry and exit for the factory, i.e. all product enters at $x = 0$ and leaves at $x = 1$, and assuming a 100 % yield, the density must satisfy

$$\begin{aligned} \frac{\partial u}{\partial t} + v(u) \frac{\partial u}{\partial x} &= 0 \\ u(x, 0) &= f(x) \\ u(0, t)v(t) &= \lambda(t) \end{aligned} \quad (2)$$

with a state equation

$$v(u) = \phi(u). \quad (3)$$

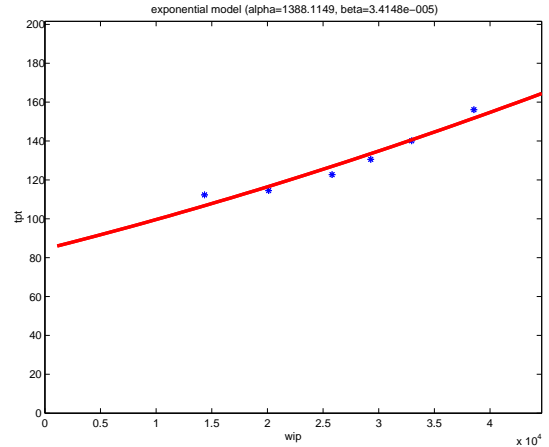


Figure 1: Throughput time versus WIP levels in steady state for a large scale discrete event simulation.

Notice that the start rate $\lambda(t)$ into the factory enters as the boundary condition for the local throughput at $x = 0$. The state equation relates the speed of the product moving through the factory to the amount of product in the factory, i.e. to WIP. The state equation (3) is based on the steady state behavior of the factory. It represents the *average* speed of a product in the factory as a function of the *average* load in the factory. Since this model does not have a time evolution of the velocity, it implicitly assumes that, whenever the load in the factory changes, the velocity will follow instantly to the new velocity given by the state equation. This is a typical characteristics of a quasi-static model also known in thermodynamics as an adiabatic model. As such, the model (Eq. 2) suffers from the same problem as the models based on clearing functions (Graves [6] and Karmarkar [9], and recent work by Asmundsson et al [4]): Adiabatic models should do well for small and slow changes of inputs into the factory but may not be as good for larger and faster changes.

In [3] we compare the PDE equation with a large scale simulation of an INTEL Corporation factory. The simulation corresponds to a full scale discrete event model of a real factory. The model contains approximately 100 machines, simulates 250 steps for a product

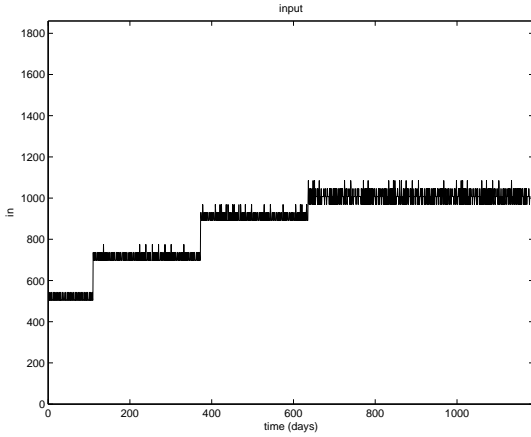


Figure 2: Start rate ramp up for the discrete event simulations

mix of 10 products. We are running this model for 6 different start rates and determine the associated average WIP levels and throughput times in steady state. This results in Figure 1. The interpolating curve in the figure follows a clearing function suggested by Asmundsson [4] and is a least square fit of the 6 data points to

$$\tau = \frac{W}{\alpha(1 - e^{-\beta W})} \quad (4)$$

with W the WIP in the factory and α and β are determined by the fit. With $v = \frac{1}{\tau}$ as steady state equation we now study the behavior of the factory to a successive ramp up of the start rate over about a 1000 days. Figure 2 shows the start rate increasing from 500 per day to a 1000 per day in 4 different plateaus. We simulate the PDE with a deterministic start rate that is constant on a plateau but follows the average start rates of the discrete event simulation. We compare the output of the discrete event simulation and the PDE in Figures 3 and 4. Note that a single discrete event simulation run takes 1 hr/year on a standard desktop computer, whereas the PDE simulation takes seconds. While the output of the discrete event simulation varies dramatically day by day, the PDE simulation is deterministic and seems to be more or less centered on the average of the output. To illustrate further how the discrete event simulation and the PDE simulation are related we smooth the output by averaging the output over 3 days in the

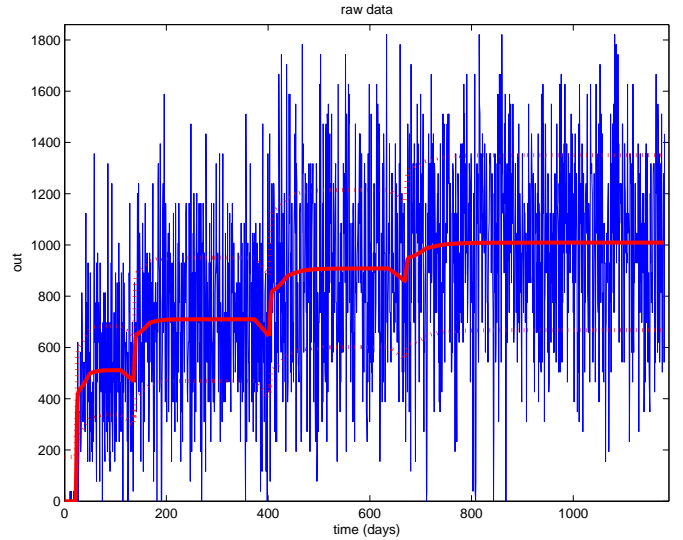


Figure 3: Discrete event simulation (blue), PDE simulation (full red), and a measure of the variance associated with the PDE simulation (dashed red): raw outputs.

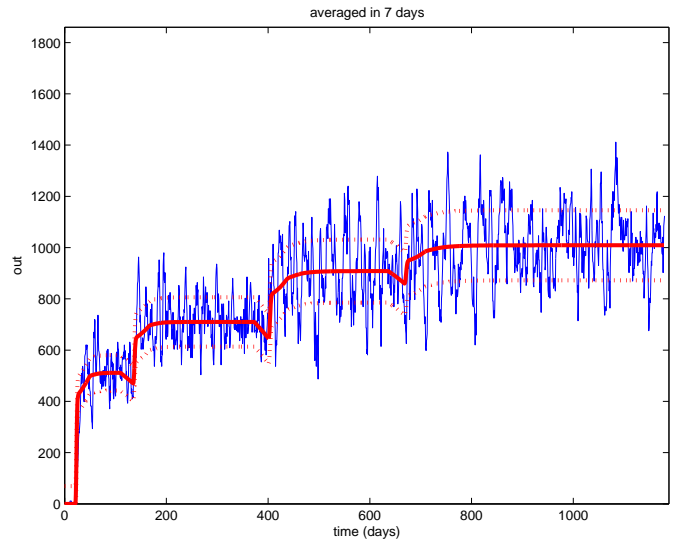


Figure 4: As Figure 3 but outputs smoothed over 7 days

future and 3 days in the past, i.e. a moving 7 day average. Figure 4 shows a quite satisfactory agreement between the averaged outs and the PDE simulation.

In addition we make an interesting observation that we most likely would not expect to find if we are doing only a pure discrete event simulation: Theoretically, for a re-entrant factory an increase in the start rate leads to an inverse response in the output, i.e. the output initially *drops* before it increases to the new level. The direct discrete event simulation (Figure 3) shows no such inverse response - if it is there it is masked in the raw outputs by the daily variation in the outputs. It is also not commonly found in the reality of the factory - due in part to the fact that operators are paid to *maintain or increase* the outputs and hence they will work hard to *speed up* WIP at the end of the production line. However, the smoothed outputs in Figure 4 indicate that, without the change of output policies resulting from operator interference, the inverse response can be found in the discrete event simulation too and it follows quite well the PDE simulation.

3 Continuous models with diffusion

Recently [1] we have developed a gas dynamic model to study the interaction between products moving through a re-entrant production line. We develop the concept of a stochastic phase velocity which dynamically updates the throughput time estimate. This leads to a diffusion term in the evolution equation for the WIP density.

The basic idea is the following. Instead of modeling the whole factory as one machine, we model a group of machines together as a machine M_m which is working at a variable speed, given (randomly) by a distribution depending on the WIP $W_m(t)$ in that group of machines as a function of time. Thus, later arrivals influence the throughput time of a given lot by influencing the speed with which the group of

machines M_m is able to process. The competition between the various steps at the reentrant machines will be governed by a set of policies such as FIFO, PUSH or PULL [12], which we assume to be given. This results in a stochastic model which is formally equivalent to a Monte Carlo discretization of a Boltzmann equation. The corresponding kinetic evolution equation for the probability density has been derived, resulting in a so called traffic flow model [7], [8] for the parts in the system. Finally, a model for the expectations of the flux through the system for large time scales via a Chapman - Enskog expansion is given by drift diffusion equation

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial F}{\partial x} &= 0 \quad x, t > 0, \\ F(x, t) &= C(t)\rho - D(t)\frac{\partial \rho}{\partial x}, \end{aligned} \quad (5)$$

$$\begin{aligned} F(0, t) &= \lambda(t), \\ \rho(x, 0) &= 0. \end{aligned}$$

Notice that this is again a quasi-static model since the drift $C(t)$ velocity is the expectation value of $\frac{1}{TPT}$ relative to the steady state distribution of the throughput times as a function of WIP in the factory. Alternatively, the drift velocity can be determined experimentally as in Figure 1.

3.1 Experimental determination of the diffusion coefficient

Figure 5 shows the paths of 920 lots as they go through an actual re-entrant INTEL factory. The underlying data set represents the times for 8 machines, when each individual lot arrived at the queue at a machine and when it left that machine. In Figure 5 we have reset all the start times to zero and plotted the time as a function of completion of the factory. Figure 6 shows the histograms of the positions at 4 different times. It clearly shows the diffusion of the δ -pulse as it moves through the factory.

In order to determine the diffusion coefficient from these data we compare the data to a

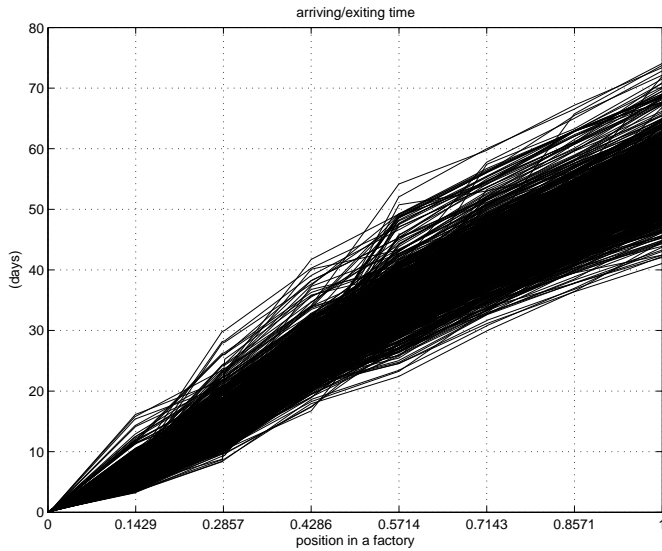


Figure 5: Paths of 920 lots through an INTEL factory

Gaussian solution of the drift diffusion equation (Eq. 5) characterized by a drift velocity v_{drift} and a variance σ^2 . We choose as an initial condition a normal distribution that we fit to the data at $t = 1$ using a maximum likelihood estimation that matches the mean μ and variance from the data. With this initial condition and zero boundary conditions at infinity we simulate (Eq. 5) and compare the calculated WIP distribution with the data at different times. The resulting WIP profile depends only on the diffusion coefficient as a free parameter. Minimizing the least square error between the simulation and the data gives us the best match for the diffusion coefficient describing the WIP distribution at a particular time. Figure 7 shows the results. The diffusion coefficient is about $0.6 \cdot 10^4 \pm 35\%$. Figure 8 shows two WIP profiles for a WIP wave going through the factory as a result of a step function increase in the influx. The sharp profile corresponds to a simulation without diffusion, the rounded one uses a diffusion coefficient of 10^{-4} .

Acknowledgments: This work was in part supported by grants from INTEL Corporation and by NSF grant DMS-0204543.

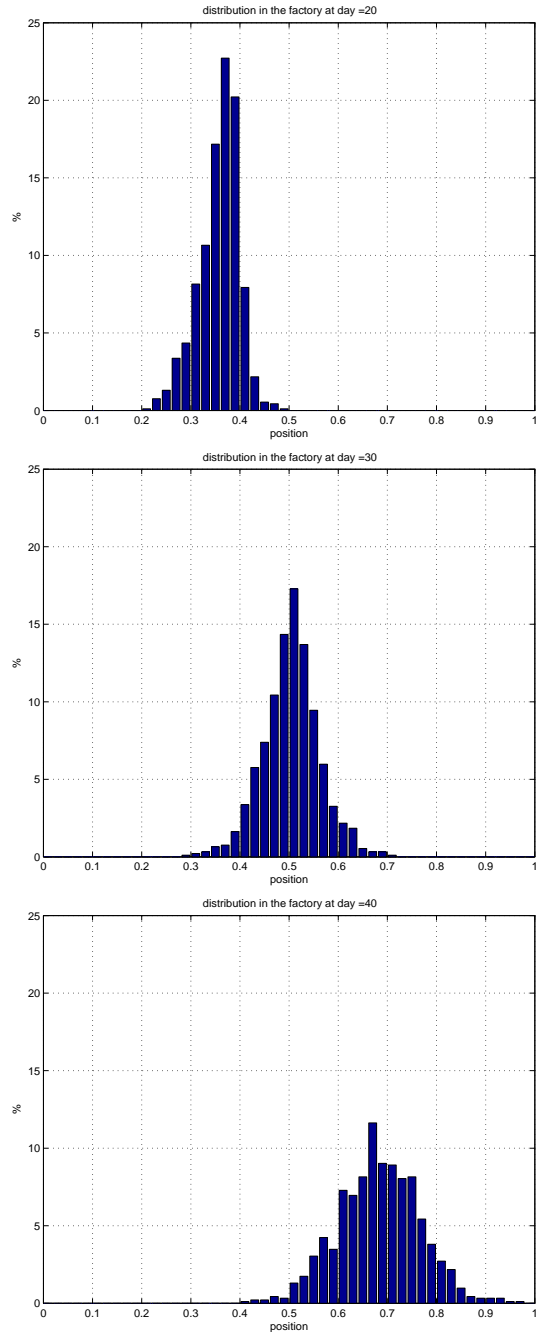


Figure 6: Histograms of positions of the lots in the factory at time $t = 20$, $t = 30$ and $t = 40$.

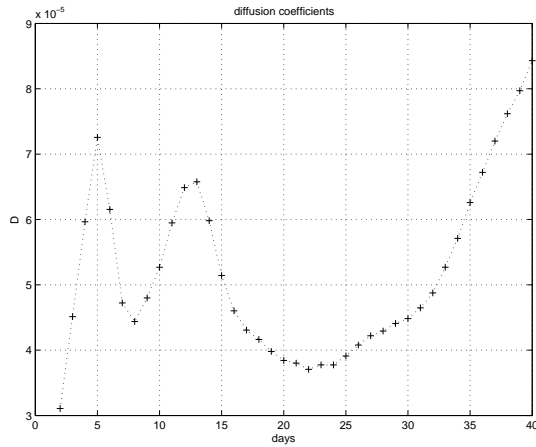


Figure 7: Diffusion coefficients that characterize the WIP distribution at various times.

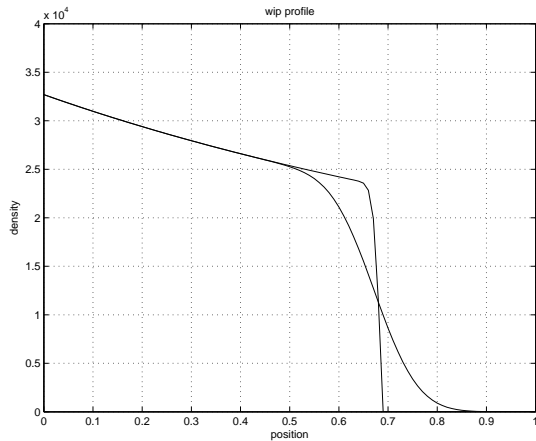


Figure 8: WIP profile with and without diffusion.

References

- [1] Dieter Armbruster, Christian Ringhofer Thermalized kinetic and fluid models for re-entrant supply chains, preprint 2003
- [2] Dieter Armbruster, Daniel Marthaler, Christian Ringhofer: Kinetic and fluid model hierarchies for supply chains, in print SIAM J. Multiscale Modeling and Simulation (10/2003).
- [3] Dieter Armbruster, Daniel Marthaler, Christian Ringhofer, Karl Kempf, Tae-Chang Jo: A continuum model for a re-entrant factory, preprint 38 pages 10/2003, in revision for Operations Research
- [4] Jakob Asmundsson, Reha Uzsoy, and Ronald L. Rardin: Compact nonlinear capacity models for supply chains: Methodology, preprint, 2002, Purdue University
- [5] Dai, J. G., Weiss, G., (1996). Stability and Instability of Fluid Models for Reentrant Lines. *Mathematics of Operations Research* **21**(1), pp.115-134.
- [6] S. C. Graves: A tactical planning model for a job shop, *Operations Research* **34** 525-533 (1985)
- [7] D. Helbing: Gas kinetic derivation of Navier Stokes like traffic equations, *Phys. Rev. E* **53**, pp. 2366-2381
- [8] D. Helbing: Traffic and related self-driven many particle systems, *Reviews of modern physics* **73**, pp. 1067-1141 (2001).
- [9] U.S. Karmarkar: Capacity loading and release planning in Work-in-Progress (WIP) and Lead-times, *J. Mfg. Oper.Mgt.*, **2**, 105 - 123 (1989)
- [10] Karl G. Kempf, Control-Oriented Approaches to Supply Chain Management in Semiconductor Manufacturing, this proceedings
- [11] A.M. Law and W.D. Kelton: *Simulation Design and Analysis*, McGraw-Hill New York, 1991
- [12] S. H. Lu, P.R. Kumar: Distributed scheduling based on due dates and buffer priorities, *IEEE Trans. Automat. Control* **36**, pp. 1406-1416 (1991).
- [13] G.F. Newell, Approximation methods for queues with application to the fixed-cycle traffic light. *SIAM Review* **7**(2) 223-240 (1965)
- [14] Wenlin Wang, D. E. Rivera, K. G. Kempf, and Kirk D. Smith, A Model Predictive Control Strategy for Supply Chain Management in Semiconductor Manufacturing under Uncertainty, this proceedings